INFORMS TutORials in Operations Research

Storytelling with Sports Analytics

Elizabeth L. Bouzarth, Benjamin C. Grannan, John M. Harris, https://orcid.org/0000-0001-8634-7985Kevin R. Hutson, Peter J. Keating

informs
http://pubsonline.informs.org/series/educ

TUTORIALS IN OPERATIONS RESEARCH
© 2021 INFORMS | ISBN 978-0-9906153-5-4
https://doi.org/10.1287/educ.2021.0230

Storytelling with Sports Analytics

Elizabeth L. Bouzarth,[a] Benjamin C. Grannan,[b] John M. Harris,[a] Kevin R. Hutson,[a] Peter J. Keating[c]
[a] Department of Mathematics, Furman University, Greenville, South Carolina 29613; [b] Department of Business and Accounting, Furman University, Greenville, South Carolina 29613; [c] ESPN, Montclair, New Jersey

Contact: liz.bouzarth@furman.edu (ELB); ben.grannan@furman.edu (BCG); john.harris@furman.edu (JMH); kevin.hutson@furman.edu, https://orcid.org/0000-0001-8634-7985 (KRH); pkeating.nj@verizon.net (PJK)

Abstract As the use of analytics grows in the sports industry, debates about the usefulness of analytical models in sports has also grown. There is no doubt that analytics have impacted the sports industry in many positive ways, but it is an evolving story as analysts seek better models of player/team performance evaluation, forecasting, and decision making. Communicating new results in these areas requires analysts to connect with organizations and fans by putting the results in context to tell a more complete story. In this work, we give examples from our own work and the work of others showing how to frame analytics within a story. At the same time, we give a brief history of the evolution in the descriptive, predictive, and prescriptive areas of sports analytics. Although this work is not meant to be exhaustive, it highlights some of the major issues that analysts face in building useful models in these areas. This paper also represents a decade-long collaboration between academics and sports writers, and we highlight some of the lessons we have learned from that collaboration.

Keywords sports analytics • storytelling • modeling • performance evaluation • strategy

1. Introduction

Jeffrey Ma and a group of fellow MIT students played a lot of blackjack in the early 1990s. They used mathematics to capitalize on betting situations, and in doing so they won a lot of money from casinos. In his book *The House Advantage: Playing the Odds to Win Big in Business*, Ma [54] describes how the team would use a single number to make decisions. It was typical practice for casinos at the time to use a small number of decks of cards at their blackjack tables and to run through the entire set of decks before reshuffling. Ma's team took advantage of the fact that when a particular table had seen a disproportionately large number of small cards, it was more likely for subsequent deals at that table to be higher cards, providing more opportunity for player blackjacks and dealer busts. If a scouted table was in this situation, a big bettor could swoop in and play the table with this knowledge. To do this, Ma's team kept count of how many face cards had been revealed at a table and how many decks remained to be used at the table. This count described the table's history and current state, it predicted future states of play, and it prescribed how to make strategic future decisions. The count told a story.

1.1. Analytics

Whereas Ma's [54] team could focus on one number to describe, predict, and prescribe actions around the blackjack table, businesses have to deal with much more complex systems. The emergence of high-powered computing, data-gathering technology, and a broad awareness of analytics has brought about many opportunities to ask interesting questions, make better decisions, and gain new insights into most business settings. The business of sports is no exception. Whereas the business side of sports has been using analytics for some time (see Lewis's [52] *Moneyball: The Art of Winning an Unfair Game*), more organizations have begun to apply analytics to player performance and decision making. Whereas in years past, in-game decisions by coaches or personnel decisions by management might have been made on "gut" feelings and hunches, today organizations employ sophisticated analytical tools to aid in these decisions. Whether the objective is to win more games or to maximize profits, sports organizations engage in the same types of operational planning as typical business organizations, finding ways to evaluate their workforce, building predictive models for forecasting, and optimizing decisions for enhanced performance. A quick glance at the growing list of attendees and speakers at sports analytics conferences in recent years is a testament to how fast this industry has adopted analytics to aid in decision making.

The use of the term analytics varies, so we adopt the taxonomy of Delen [24] (also promoted by INFORMS [40]) to classify analytics into three categories: descriptive, predictive, and prescriptive. Descriptive analytics uses mathematical and statistical techniques to describe the current state of a process and what has happened in the past to lead up to that state. These techniques might include reporting on some evaluative metric or showing some exploratory data analysis or visualizations to describe the current state of the operations. Predictive analytics tries to forecast what will happen in the future. Tools used in this area to help quantify the likelihood of future events include regression analysis, time series models, and machine learning techniques. Prescriptive analytics attempts to identify optimal decisions given objectives and current states. Optimization tools such as linear, nonlinear, and integer programming, along with simulation models, are employed in this area to aid decision makers in prescribing courses of action for their organizations. In this article, we will look at examples from the sports world in each of these areas, showing how analytics and the principles of storytelling can be used together to craft impactful stories.

1.2. Storytelling

As important as it is for companies to use good analytical approaches, it is equally important for them to be able to communicate the results of the analysis in an effective manner. As few audiences enjoy presentations consisting strictly of numbers and statistics, this communication part is crucial. To make lasting impressions, analysts have turned to incorporating their messages into story narratives. Knaflic [48] blends classical storytelling principles (understanding the audience, focusing the audience's attention, incorporating the three acts of beginning, middle, and end) with the use of visualization and analytics tools to craft an impactful and inspiring narrative. Audiences have a way of engaging with stories so that they emerge from the story with new positions or perspectives. Rein et al. [73] note that narratives incorporate and emphasize certain positions and perspectives shaping events rather than simply recalling facts. Audiences mentally transport themselves to the world of the story and are more likely to detach from previous attitudes and beliefs than if directly confronted with an argument.

In other words, analytics can have a great impact when presented within a story framework.

Drama is a natural element of sports: there are ready-made conflicts, competing heroes, and natural cliff-hangers. Seasons and games are themselves three-act plays with beginnings, middles, and ends. Compelling plot lines, personalities, and settings have provided a depth, emotion, and complexity that sports writers tap into when writing stories. Furthermore, these plot lines, personalities, and settings tend to show up year after year. Knaflic [48] notes that there is a power in repetition to help an audience remember. Sports fans relate to the enduring stories of the unlikely rise of the underdog, the clutch comeback, or the notion of fair play. We enjoy arguing around water coolers and on talk shows about bad calls and about players who are either overrated or the greatest of all time. The market for these discussions is lively, and it drives the creation of corresponding journalism. Analytics can play a role in this discussion as well, as it provides verification, justification, and the basis of criticism for the decisions made in the sports industry.

In what follows, we adopt the basic storytelling framework described in Robert McKee's [59] book *Story*. Building on Aristotle's three act concept of story, McKee [59] describes the basic composition of a story as having five components:

1. The inciting incident—the question that the story is trying to address and the inspiration for a potential space of possible solutions;
2. Progressive complications—the impediments to resolving the question, including the constraints of setting and character;
3. Crisis—the motivating event that influences the main character to make decisions toward resolving the question;
4. Climax—the actions of the main character to resolve; and
5. Resolution.

Professionals in operations research are familiar with these story components because they also serve as the basic building blocks of search algorithm design. The story/design process starts with a question to answer, a clear set of goals for the storyteller. There are numerous possibilities with which to begin the story, and the author must choose an initial state for the hero. The author chooses a direction from an infinite number of possible plot lines, and the hero begins the quest, guided by an objective and constrained by the boundaries of the setting and characters. The hero achieves new states, and an iterative process of making decisions about story directions commences until a resolution is found. *Story is the tale of assessing and evaluating states, searching and decision making, and resolving the problem.*

The work that we describe here is built on a decade-long collaboration between academics and sports writers. We address the three areas of analytics, and we give examples (from our own work and the work of others) that show how to present the work in terms of McKee's [59] storytelling framework. As is often the case with presentations of analytical results, we have removed much of the underlying mathematics in favor of the story. We do, however, provide references for the interested reader. Sports is an evolving story in its own right, and we identify examples of how the story of sports is changing based on analytics, providing the reader with a breadth of examples of analytics work in the area of sports. Finally, we end this paper by discussing lessons we have learned through this collaboration.

2. Descriptive Analytics: Performance Evaluation

Economists track numbers like the consumer price index and gross domestic product (U.S. Bureau of Labor Statistics [86]). Advertisers track website clicks and the cost of thousands of views (Investopedia [41]). Meteorologists track the air quality index (AirNow [1]), and healthcare professionals track a person's body mass index (NIH National Heart, Lung, and Blood Institute [68]). As was the case with the MIT blackjack team, the motivation here is to use with a single number to tell the story about the state of a process or system.

Much of the research in and around the sports industry has sought to find metrics that are useful in evaluating players, teams, decisions, etc. If a number could be found that measures the performance of every player (or coach, manager, etc.) and team in a sport, it could aid in salary negotiations and in making decisions about playing time, team composition, and playoff determination. However, it has proven difficult to find one metric of performance that incorporates all desired factors, and measures often tell an incomplete story for several reasons. It is difficult to find evaluative performance metrics in every sport that eliminate the influence of luck, that extract one player's contribution from the team contribution, and that incorporate comparative inconsistencies such as playing fields and opponent matchups. For some sports like baseball, the individual parts of the game (pitching, batting, fielding) make it possible to isolate metrics to each player (wins against replacement, on-base plus slugging, ultimate zone rating, walks and hits per inning pitched, etc.; see Meoli [60]), but these types of metrics have an evolutionary history, as analysts have tried to remove the influence of luck and park effects. In low-scoring, team-based sports like soccer, it can be difficult to ascertain what game metrics best represent an individual player's contribution to a goal or team win.

Team quality is also difficult to quantify for many of the same reasons. Debates take place every year regarding which teams should be included in playoffs or end-of-the-year tournaments. A team's chances of making end-of-season playoff tournaments or even winning those tournaments is highly dependent on the way teams are partitioned into groups, conferences, and divisions, and how teams are seeded in those tournaments. The fairness of these groupings and the resulting team schedules significantly impact the way team performance is evaluated. Metrics often used for justifying teams' inclusion or exclusion from the these tournaments are based on winning percentage, strength of schedule (SOS), margin of victories, the "eye test," etc. (see Langville and Meyer [50] for an overview).

Analytics provides a common frame of reference for data generated in various settings and contexts (different eras, stadiums, styles of play, etc.). Pace of play in basketball is a good example of this. Even though opposing teams in a particular game will have nearly identical possession numbers for that game, there is wide variance in this area in the long run. To understand and describe team pace accurately, it is important to adjust statistics for tempo. Frank McGuire [58], a coach at the University of North Carolina, had this insight as early as 1959, when he published a book chapter called "The Importance of Possession Statistics." Researchers such as Dean Oliver [70] and John Hollinger [37] began popularizing the idea around 20 years ago. One example of this sort of thing is the 2019 University of Virginia men's basketball team. The team scored 2,299 points in 32 men's college basketball games, for an average of 71.8 points per game, which ranked them 196th in the National Collegiate Athletic Association (NCAA; Pomeroy [71]). But the Cavaliers played at a pace of just 61.6 possessions per game, the slowest tempo of any Division I team. Scoring 116.6 points per 100 possessions, Virginia actually had the second most efficient offense in the country—and they won the national championship that year.

We now look at two examples highlighting the complexities of performance evaluation within the storytelling framework. The first originates from a series of articles written by Peter Keating [44] on why the Bowl Championship Series (BCS) college football national championship selection system was flawed. The second comes from the work of Kirk Goldsberry [32, 33] showing how player evaluation in the National Basketball Association (NBA) changed as a result of data analytics.

2.1. Rating Systems: Evaluating Team Performance

2.1.1. Inciting Incident: How Do We Evaluate Team Quality? Measuring the quality of teams within a sports league is often made difficult by a number of factors: lack of head-to-head matchups between teams, injuries to star players, and variation in schedules. Furthermore, it usually does not take long in any season for it to be the case where team A has beaten team B, which has beaten team C, which has beaten team A, making it even difficult to produce comparative rankings. As Arrow's [2] impossibility theorem attests, no ranking system will satisfy all of the conditions that a nonsubjective ranking would aspire to satisfy.

Many mathematical methods have emerged to rate/rank team quality (see Langville and Meyer [50] for an exhaustive treatment). Whereas Coleman [21] uses integer programming techniques to rank teams to minimize the number of teams ranked higher than teams who defeated them, others (Callaghan et al. [19], Colley [23], Kvam and Sokol [49], Massey [55]) build a weighted network of games based on the results of the head-to-head matchups and use matrix manipulations (solving systems of equations, Markov chains, etc.) in this network environment to establish a rating for each team that includes a strength of schedule component. These mathematical rating and ranking methods help combat the lack of head-to-head results between teams without incorporating a subjective bias as the game network provides paths to help compare teams.

2.1.2. Progressive Complications: With Limited Games and Subjective Polls, College Football Seems Especially Problematic to Evaluate Team Quality. Evaluating college football teams, in particular, is challenging. They have limited schedules that are concentrated within conferences, and this makes it hard to identify the true quality of each team. In the late 1990s and early 2000s, the BCS emerged to address the difficulties in the selection of a national champion. Prior to that time, there was no single championship game, and champions were determined by which team or teams finished at the top of subjective sports writers and coaches polls. The BCS selected two teams to participate in a single championship game based on several components, including two polls (the Harris Interactive College Football Poll and the USA Today Coaches Poll) and a consensus of six mathematical ranking systems (Wright [87]). These ranking systems were restricted in the types of inputs that could be considered in ranking teams such as "quality" wins and losses and strength of schedule. To discourage teams from running up the score in games, these methods were not allowed to use margin of victory in the computations.

2.1.3. Crisis: The BCS Formula Did Not Allow for Margin of Victory to Be Included in the Calculation. The exclusion of margin of victory from the ratings calculations made it difficult to differentiate between teams with the same win–loss records, but where one team won close games and the other won blowouts. The Massey method (Langville and Meyer [50]), also known as the simple rating system (SRS), includes margin of victory within the calculation, although Massey had to modify this system to be included in the BCS calculation. The SRS method produces a rating for each team that gives an indication of expected margin of victory of one team over another by comparing the two teams' SRS scores. The basic tenet of the SRS rating system is that each team is given a rating equal to its average margin of victory plus an average rating of its opponents (a strength of schedule component). Further constraining the system to require that all ratings sum to zero ensures that a team's SRS score can be interpreted as that team's margin of victory over an average competitor. For example, an SRS score of 10 means that a team is 10 points better than an average team on a neutral field.

2.1.4. Climax: In 2013, Notre Dame Is Chosen by the BCS to Play Alabama in the Championship Game. In 2013, the Massey method ranked Alabama at the top with a rating of 35.4 (meaning the Crimson Tide was 35 points better than an average team on a neutral field). Florida was ranked second with a rating of 28, whereas Notre Dame was in a dead heat with Oregon for third and fourth, nine points behind Alabama's rating. Florida had one loss, whereas Notre Dame was undefeated. Florida had the best defense in the BCS and had played the toughest schedule in the country, whereas Notre Dame had several victories of smaller amounts against mediocre competition. However, because defensive prowess could not be part of the computerized ratings, Notre Dame was chosen over Florida. Notre Dame ended up losing to Alabama in the championship game 42–14, the second largest defeat in the BCS era. Who knows how Florida or even Oregon would have fared in the game, but the game added fuel to the fire around the debate of an expanded playoff system.

2.1.5. Resolution: The BCS Is Replaced with the College Football Playoff. The two teams selected by the BCS often led to much debate about the BCS formula. Furthermore, there was a tension between mathematical methods, which reduced subjective bias but had only limited inputs, and human perception, which had many more inputs but also had bias. In response to pressure for transparency in team selection and to frustration with mathematical rankings not aligning with the eye test, the College Football Playoff replaced the BCS system in 2016. The playoff expanded the field to four teams instead of just two, and selection is now made by a committee. The committee takes into account championships, unavailability of key players, strength of schedule, head-to-head competitions, and comparative match outcomes (College Football Playoff [22]). The committee is not influenced by polls where initial rankings are established before competition has occurred, such as sports writers polls that publish preseason rankings. The committee has a voting process that results in identifying the four teams and their seeding for the tournament. This only appeased fans temporarily as calls persist to expand the playoff to eight or more teams.

In some sense, the BCS system was the perfect solution to choosing participants for a national championship game. It married the objective mathematical network-based systems with the subjective, voting-based systems, which could incorporate factors that could not be quantified mathematically. Its failure signals the need for analytical rating systems that can incorporate more metrics and eliminate the subjective human bias in these decisions.

2.2. Changes in Strategy Impact Player Evaluation

We can use statistics to rank athletes or teams, or to decide how much to pay players, but maybe analytics' biggest impact over the past generation has been discovering which kinds of performance are most valuable and then leading teams to focus on acquiring players who can deliver that performance, and to tilt their strategies toward getting it.

In baseball, this work began almost as soon as interactive terminals became available for mainframe computers. In 1970, for example, Harlan and Eldon Mills, brothers who were also both military pilots and data analysts, published *Player Win Averages* (Mills and Mills [61]), a book that contained all the elements of win probability. The Mills brothers leased an IBM 1620 and fed it Fortran-coded punch cards to simulate thousands of baseball games. They calculated the likely result of various combinations of factors, such as the inning, runners at each base, and number of outs. Then they rated Major League Baseball (MLB) players from the 1969 season based on the cumulative value of the changes those players produced. Soon afterward, researchers began using the estimated value of game states to calculate the utility of strategies that require trading one bundle of situational variables for another—and the results were remarkably consistent. "Every mathematical analysis I've seen shows that the intentional walk is almost always a bad play" (Thorn and Palmer [83]), pioneering baseball analyst Pete Palmer said in 1983. "Stolen bases are only marginally useful, and the sacrifice bunt is a relatively useless vestige of the deadball era when they didn't pinch-hit for pitchers."

Similar studies of football took longer to develop, where far more variables combine to create particular game states than in baseball. But in the 2000s, play-by-play analysis of National Football League (NFL) games revealed something just as revolutionary: passing the ball is more efficient than rushing, by a fairly huge margin. "In fact, for most of the field, the average value of a run is essentially zero or negative," analyst Brian Burke [18] wrote in 2009. And NFL passing has increased in 11 of the past 15 years, with 2020's total of 240.2 yards per team per game the second highest total in league history (Pro Football Reference [72]).

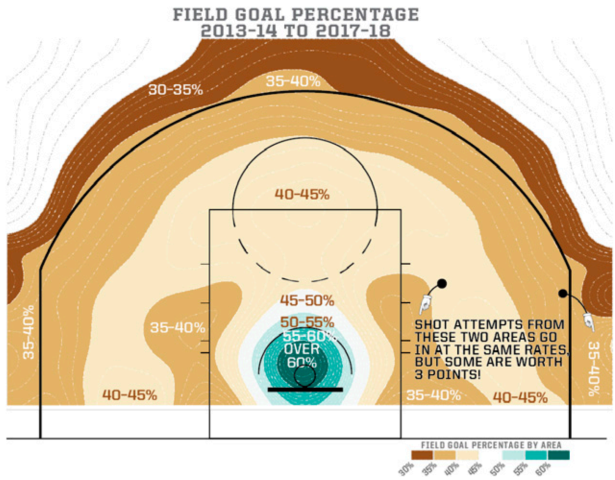
2.2.1. Inciting Incident: How Do We Evaluate Players in a Team-Based Sport? It is difficult to extract an individual player's performance contribution in a team-based sport. In basketball, the main roles that a player contributes to are shooting, assisting other players, defense, and rebounding. Although there may be measures of each of these individual contributions that a player can make, how each of these contribute to a win is hard to unpack. Role players on one team might be superstars on another based on the offensive schemes that a coach runs. Although superstars in basketball are normally those players who shoot well, a player's main contribution to a team win might be tied to the other main roles, and one player leading the league in scoring might be dependent on other players creating the conditions in which the first player can score. This type of conditional evaluation contributes to the difficulty in rating a player's performance.

Metrics to rate players in each of their roles in the NBA have evolved past traditional metrics like points per game, field goal percentage, assists per game, rebounds per game, steals, and blocked shots. After adjusting for pace of play, some metrics, such as the player efficiency rating (see Hollinger [38]), use a combination of traditional measures to form a composite rating. Recognizing that a player's performance is tied to the teammates playing on the court at the same time, other metrics, such as adjusted plus-minus systems (see Ilardi and Barzilai [39], Rosenbaum [75]), look to evaluate how well the team plays together when a player is in the game, whether the point differential is positive or negative during that time. For example, box plus/minus uses box score data to estimate a player's contribution to his team while he is on the court (Myers [65]), and the value over replacement player compares that performance to that which a replacement-level (or minimally competent) player would provide. Both are now available on the website Basketball Reference [8]. Deshpande and Jansen [25] use a Bayesian linear regression model and focus on win probability and game context (i.e., whether the score is a blowout or close game) to evaluate players. The use of motion tracking systems has allowed for defensive metrics to become more advanced to track player-specific opponent shooting percentage and "openness" of a shot (Silver [79]). As these metrics evolve, there are still issues related to finding player metrics that measure player-team chemistry, fit for certain styles of offenses/defenses, and combining relevant metrics into a composite score.

2.2.2. Progressive Complications: Rule Changes to the Game Bring About Strategy Shifts for Offenses. When the NBA introduced the three-point line in the 1979–80 season, most teams saw long-range shooting as a gimmick, and players took just 3% of their shots from behind the arc that year (Basketball Reference [4]). Michael Jordan, the league's deadliest shooter, several-time most valuable player, and the first NBA player to sign a contract for more than $20 million, attempted only 7% of his field goals from three-point range in a career that spanned from 1984 to 2003 (Basketball Reference [7]). Since the Jordan era of basketball, the NBA has seen a shift in strategy related to the three-point line. From 2012 to 2018, Stephen Curry broke the single-season three-point record four times while winning two Most Valuable Player Awards and leading the Golden State Warriors to three league championships. From 2016 to 2020, the Houston Rockets, whose general manager at the time was Daryl Morey, a longtime leader in basketball analytics, set new standards for three-point shooting by a team. Midrange jumpers (shots taken from anywhere between 10 feet from the basket out to the three-point line) had represented almost 40% of NBA shots 20 years ago (Basketball Reference [5]), but that proportion dropped to under 17% in the current season (Basketball Reference [6]).

2.2.3. Crisis: Descriptive Analytics Provided an Impetus for the Strategy Change. Better descriptive analytics in basketball has led to major changes in how the sport is played. In the 2010s, statistical analysis (Goldsberry [33]) revealed that although NBA players' average shooting percentage declines with distance, it does so gently. Sixty percent of shots taken near the basket are made, but this field goal percentage decreases to less than 40% as the radius from the basket expands beyond the three-point arc. As Goldsberry [33], a pioneer in shot mapping, asked: "If it's true that three-point shots go in 36 percent of the time and 10-foot shots go in just 40 percent of the time, then why are we assigning 50 percent more value to shots from beyond that magical little arc?"

Figure 1. Heat map of field goal percentage in the NBA from 2013 to 2018.



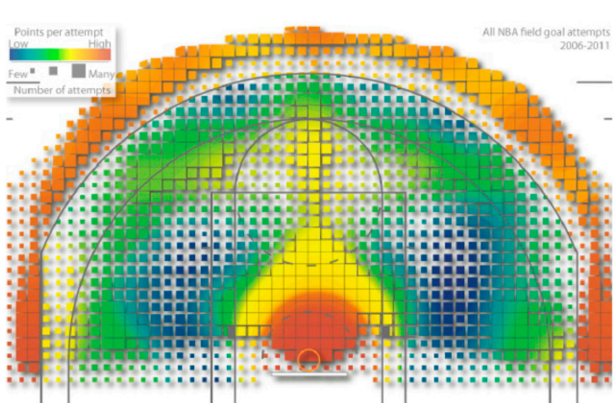
Source. Goldsberry [33]. Copyright ©2021, INFORMS. Republished with permission.

As Figure 1 (from Goldsberry's [33] work) shows, a majority of shots in the NBA are now taken from either right underneath the basket or from behind the three-point arc. The percentage of shots taken from in between these two areas has decreased significantly compared with previous generations.

2.2.4. Climax: Where Should Shots Be Taken from if a Team Wants to Maximize Its Points per Possession? Analysts, including Goldsberry [33], combined the field goal percentage at different areas of the court with the amount of points scored from each of those areas to come up with an expected points scored on a shot taken from each location. The result was that shots taken very close to the basket and shots taken from outside of the three-point arc all had an expected point value greater than one. Meanwhile, shots taken between these areas had expected points ranging from 0.8 to 0.9; see Figure 2 (from Goldsberry [32]).

2.2.5. Resolution: Player Evaluation Is Tied to Team Strategy. Zwerling [88] discusses that one of the dominant game strategies in the NBA is to run a high pick-and-roll offense. In this offense, the dribbler brings the defender into another player setting a screen

Figure 2. Heat map of points per field goal attempt in the NBA from 2006 to 2011.



Source. Goldsberry [32]. Copyright ©2021, INFORMS. Republished with permission.

near the three-point line. If the two defenders switch who they are guarding, it usually results in a mismatch where a smaller player is guarding a taller player or a slower player is guarding a faster player. This allows either the faster player to drive toward the basket, which usually results in either a shot near the basket or a pass to an open three-point shooter whose defender came off to try to help defend against the driver. This type of offense sets up a scenario where the team is able to maximize the expected points of its possession as determined by the aforementioned analysis. It also de-emphasizes the midrange jumper. As Zwerling [88] points out, teams wishing to take advantage of the analytics then seek players in the draft and through trades who can help them achieve success in this offense scheme and increases their value in the draft and in the marketplace.

3. Predictive Analytics

As we saw in the last example, descriptive analytics not only describes the current state of a process or system, but those metrics can also be used as a basis for making future strategic decisions. Using analytics derived from past performance to predict future performance is well studied within the area of business forecasting. In the business of sports, examples include using performance in college to predict performance in the pros (draft analytics) and predicting revenue from merchandising and concessions based on past games or past attendance levels, weather conditions, marketing campaigns, etc. (Harrison and Bukstein [36]). There is also a booming business around predicting future sports performance as the casual fan engages not only in betting outlets in the gaming industry but also in recreational outlets in the daily fantasy sports industry (Nover [69]).

Anyone who takes part in these industries knows the difficulty in predicting future events in complex systems. There are several things to consider when making predictions about performance. Most performance evaluations are comprised of a mixture of skill and luck. The most skilled team does not always win the game. Several analyses, see Mauboussin [56], for example, show that as skill in league increases, luck plays a larger role in determining outcomes, making outcomes harder to predict. From a player performance perspective, there is evidence from Bradbury [13] that suggests skill level changes as the player ages. One must also take into account the context with which past data have been generated for a team or player and the context in which future data will be generated. For example, to help determine whether a high school player will succeed at the college level, one must take into account the level of competition as well as the system in which the high school basketball player participated. Furthermore, opponent style of play may have a big effect on the outcome of a future game as when ball-control offenses play against fast-break offenses. There is also a temptation to make predictions in scenarios where there are only small sample sizes and limited information. For example, teams might ask, how much should recent data be weighed? Classic Bayesian analysis says conditional information does not always have large effects on prediction, and regression to the mean employs the same concept. It is common to see drastic evaluation changes based on recent information only to see a reversion to the mean—see the sophomore slump, the *Sports Illustrated* jinx, and the plethora of coaches and players who have seen high pay raises in one year only to be followed by mediocre seasons (Dilipkumar [26]).

In this section we look at two storytelling examples. The first is from our own work trying to predict NCAA tournament upsets in college basketball. It looks at a simple example of how to use matchups to aid in the prediction. The second example is from baseball, highlighting the need to consider what role luck plays in baseball statistics.

3.1. The Underdog Story

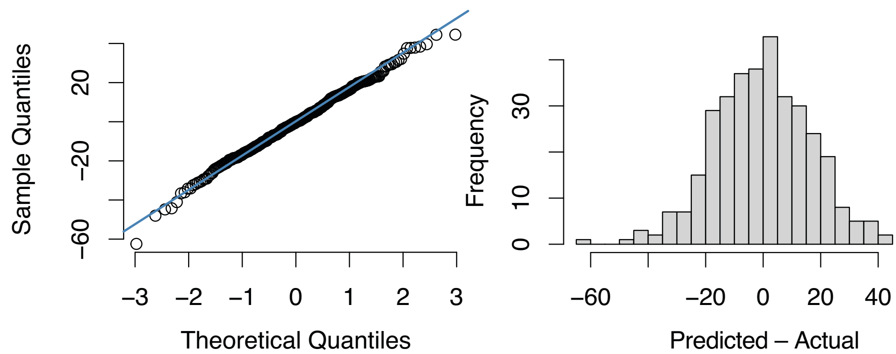
The story of the underdog, like David and Goliath, is an ancient story of a seemingly overmatched opponent that finds a way to win a competition against all odds. There have been many memorable examples of underdog victories in sports: the U.S. National Hockey Team's upset of Russia in the 1980 Olympics, the Miracle 1969 New York Mets, and 16-seeded UMBC's upset of 1-seeded Virginia in the 2018 NCAA Men's Basketball Tournament. The last of these was pretty shocking, given that most pundits argued that this would never occur. Of all of the years that the tournament has been played, this was the only time a 16-seed had won a game.

3.1.1. Inciting Incident: Which Team Is Going to Pull off the Big Upset This Year in the NCAA Tournament? Or, How Do I Fill Out My Bracket for the Office Pool? The NCAA Division I Men's Basketball Tournament, also known as March Madness, is a great spectacle in sport and it inspires people across the United States to try to predict the outcomes of the games (Lunardi and Smale [53]). There have been a variety of mathematical approaches to predicting outcomes in the tournament (see BracketOdds [12], Chartier et al. [20], Dutta et al. [27], Khatibi et al. [47], Kvam and Sokol [49]). Since 2013, the authors (see Brenner and Keating [14, 15, 16], Keating [45, 46]) have tried to predict upsets in the NCAA Men's Basketball Tournament between so-called Giants and Killers. The tournament field is divided into four regions, each with 16 teams seeded 1 through 16. The tournament selection committee (National Collegiate Athletic Association [66]) uses different metrics of team quality to seed teams, but generally higher-seeded teams (e.g., 1-seeds) are top teams and lower-seeded teams (e.g., 16-seeds) are conference tournament winners from low- to midtier conferences or teams from power conferences with mediocre records. For our analysis, when a game takes place between two teams whose seed difference is at least five, the higher-seeded team is labeled the *Giant* and the lower-seeded team is labeled the *Killer*. For instance, in 2017, when the 7-seed South Carolina upset the 2-seed Duke, we labeled South Carolina a *Giant Killer* and Duke a *Slain Giant*. The goal of our work has been to determine the characteristics of Killers that make them perform better than normal and to determine the characteristics of Giants that make them perform worse than normal in a tournament setting.

3.1.2. Progressive Complications: Rating Systems Are Good Predictors but Contain a Lot of Variation. In our analysis, we use the aforementioned simple rating system to give a rating for each team. We scale each game's statistics by possessions so that comparisons between games are consistent, and a team's SRS score in our analysis gives an indication of the expected margin of victory against an average competitor per 100 possessions. Additionally, the difference between two teams' ratings provides an expected point differential per 100 possessions. The difference in SRS ratings on its own does a good job of predicting winners in the NCAA tournament, as seen in Figure 3. The histogram is roughly normally distributed, centered very near zero with a large standard deviation (approximately 12). Other studies (Minton [62]) have shown a similar distribution with high deviation when comparing actual results to the Vegas line. The high variation in the SRS prediction led us to see whether we could do better by incorporating matchup information.

3.1.3. Crisis: The 2014 NCAA Tournament Had Plenty of Upsets but for Different Reasons. Prior to the 2015 and 2016 seasons, we employed a regression analysis approach in trying to determine a measure, which we called *secret sauce*, to determine how much better or worse a Giant or Killer would play in the tournament. We then used a logistic regression model to convert the difference in SRS scores, fudged by this secret sauce, into a probability of an upset. However, in the 2014 tournament, there were a ton of upsets. In fact, the national title game was between an 8-seed Kentucky against a 7-seed University of Connecticut. As we analyzed the results of the tournament, it became clear that the reasons 11-seed Dayton upset 6-seed Ohio State or 8-seed Kentucky upset 1-seed Wisconsin were quite a bit different from the reasons 14-seed Mercer upset 3-seed Duke or 12-seed Stephen F. Austin upset 5-seed Virginia Commonwealth University. We needed a way to move from a one-size-fits-all regression

Figure 3. (Color online) The quantile–quantile plot (left) and histogram of the difference between the predicted score of a game (using the difference between the Giant's SRS and Killer's SRS) and the actual score in the game (right).



model to a model that allowed particular matchup characteristics to contribute to the upset probability calculation.

3.1.4. Climax: Can We Use the Specific Matchup Characteristics to Predict Games? To enhance our predictive powers, we performed a *k*-means clustering on the Giant–Killer tournament games from 2007 to 2016. We separated the Giants and the Killers each into four clusters based on the variables that we found were significant in helping to predict upsets in our regression analysis. For the Giants, the variables were offensive rebound percentage, opponent offensive rebound percentage, and opponent turnover percentage. For the Killers, the clusters were based on pace, opponent effective field goal percentage, offensive rebound percentage, steals, and tendency to shoot three-point shots. The results of the clustering analysis produced four Giant cluster centers having the following characteristics:

- Giant Cluster 1: Below average in offensive rebounding, average in terms of defensive rebounding, and weak at causing turnovers
- Giant Cluster 2: Strong in offensive rebounding, average in defensive rebounding, and weak at causing turnovers
- Giant Cluster 3: Slightly above average in offensive rebounds, noticeably below average in defensive rebounding, and very highly above average in creating turnovers
- Giant Cluster 4: Average at offensive rebounds, above average at defensive rebounds, and below average at causing turnovers

For the Killers, we found the following cluster center breakdown:

- Killer Cluster 1: Play at an above-average pace, below average at keeping opponent's effective field goal percentage down, below average at offensive rebounding, average at creating steals, and below-average tendency to shoot three-pointers relative to other Killers
- Killer Cluster 2: Play at an above-average pace, above average at keeping opponent's effective field goal percentage down, very good at offensive rebounds, above average at creating steals, and way below-average tendency to shoot three-pointers
- Killer Cluster 3: Play at a below-average pace, below average at keeping opponent's effective field goal percentage down, not great offensive rebounders, above average at creating steals, and way above-average tendency to shoot three-pointers
- Killer Cluster 4: Play at a very slow pace, way above average at keeping opponent's effective field goal percentage down, strongly above-average offensive rebounders, not good at creating steals, and below-average tendency to shoot three-pointers

Table 1. Names and examples of Giant and Killer clusters.

Cluster	Giant name	Example teams	Killer name	Example teams
1	Power Giants	Baylor, UNC	Generic Killers	Xavier, Harvard
2	Gambling Giants	Villanova, Gonzaga	Slow Killers	Valparaiso
3	Pack-Line Giants	Virginia, Wisconsin	Perimeter Killers	Iona
4	Generic Giants	Oklahoma	High-Possession Killers	St. Mary's, Wichita State

Note. UNC, University of North Carolina.

As the story started to take shape around the analytics, it became increasingly obvious that the characters (clusters) were faceless. So, we attached names and classic teams to each category, as seen in Table 1. The story comes alive because now the fan has a character to attach to each of the players in the story and can recall upsets that occurred in each category.

3.1.5. Resolution: Like the MIT Blackjack Team, if You See This Matchup, Bring in the Big Money. Table 2 shows the upset percentages by cluster, where you can see that 23% of Giant–Killer tournament matchups from 2007–2016 were upsets. There are real stories embedded in the table. Cluster 1 Giants are by far the weakest Giants. They get upset 35% of the time. In fact, when facing Cluster 2 Killers, Cluster 1 Giants actually lose 55% of the time. The aforementioned 2014 upsets of Dayton over Ohio State and Kentucky over Wisconsin both were Cluster 2 Killer versus Cluster 1 Giant matchups. Cluster 2 and 4 Giants are the strongest Giants, losing just 18% of the time. Both clusters are most vulnerable, however, to Cluster 4 Killers. Cluster 3 Giants, with their decreased ability to rebound on the defensive end and susceptibility to turnovers, are also vulnerable to Cluster 4 Killers, being upset one in every three matchups.

If one is looking to predict upsets in the tournament, a good bet is to bet on a Cluster 2 Killer beating a Cluster 1 Giant. However, the NCAA selection committee has not given us the Giant Cluster 1 versus Killer Cluster 2 matchup in the first round very often recently, perhaps because of increased use of analytics on their part, but when they do, it tends to pay off. In the last three tournaments (2017–2019), the matchup has appeared four times, three of which turned out to be upsets:

1. (2017) Six-seed Maryland versus 11-seed Xavier (result: Xavier by 11)
2. (2017) One-seed Villanova versus 8-seed Wisconsin (result: Wisconsin by 3)
3. (2018) Three-seed Michigan versus 14-seed Montana (result: Michigan by 14)
4. (2019) Four-seed Kansas State versus 13-seed UC Irvine (result: UC Irvine by 6)

3.2. Separating Skill and Luck

Determining which components of performance are sustainable is key to developing predictive analytics. In his book *The Success Equation*, Mauboussin [56] discusses the roles that skill and

Table 2. Percentages of Giant–Killer tournament games from 2007 to 2016 that are upsets, broken down by clusters.

	Killer 1	Killer 2	Killer 3	Killer 4	Overall giant upset %
Giant 1	28	55	33	29	35
Giant 2	12	6	18	29	18
Giant 3	13	16	20	33	20
Giant 4	8	19	18	26	18
Overall killer upset %	15	24	22	28	23

Note. All values are percentages.

luck play in predictive analytics. He examines the idea of the persistence of a performance evaluation and measures this with the correlation coefficient r. He indicates that evaluations with high year-to-year correlations are generally consistent with skill, whereas low correlations are more indicative of luck. Thus, the higher the correlation, the better chance that the evaluation will be able to predict future performance. Even the simple steps of separating previously aggregated statistics and recognizing regression to the mean can produce powerful predictive analytics. For example, beginning in the early 2000s, research showed that although some NFL players and teams can persistently force fumbles, fumble recovery is essentially a random act. As Aaron Schatz [77], editor-in-chief of *Football Outsiders*, has written: "Stripping the ball is a skill. Holding onto the ball is a skill. Pouncing on the ball as it is bouncing all over the place is not a skill. There is no correlation whatsoever between the percentage of fumbles recovered by a team and the percentage they recover in the next year." For example, in 2020, the Carolina Panthers led the NFL by recovering 65.7% of the fumbles committed (by both the Panthers and their opponents) in their games. But in 2019, their recovery rate was just 43.9% (ranking 25th), whereas in 2018, it was 57.1% (ranking 5th; Team Rankings [82]). Teams or fantasy players relying on turnovers to be a consistent metric are in for a rude awakening.

Sometimes the task of separating skill and luck in performance requires challenging powerful assumptions. Nowhere is this more evident than in baseball, where beliefs regarding player evaluation tend to have sticking power. Millions of dollars are invested based on the promise/hope of a player's future performance, and so the desire to accurately evaluate players is understandable.

3.2.1. Inciting Incident: By What Measures Should We Evaluate Pitchers? For more than a century, front offices and fans alike have judged players on the basis of scoreboard-driven statistics. For pitchers, in particular, lots of attention has been given to total wins and earned run average (ERA).

3.2.2. Progressive Complications: Measures Are Often a Function of the Data Gathered. Baseball analytics, like analytics for most sports, can only be as sophisticated as the data gathered. Historical evaluations of pitchers centered around wins and ERA because those data were available. With Bill James [42] and others breaking from traditional metrics of evaluation, data became increasingly more available.

3.2.3. Crisis: Wins and ERA Can Be Relatively Inconsistent from Year to Year. As early as the 1930s, observant fans noticed an unexpected amount of variation in the lists of top win earners among pitchers. For instance, if a fan examined the top 10 win earners for each season between 1930 and 1934, that fan would have found 35 different pitchers who appeared on at least one of those lists. Similar variability has appeared in top ERA lists, and the phenomenon continues to this day. From 2015 to 2019, there were 37 different pitchers who appeared on at least one top-10 ERA list. What is the cause of this variability? Are pitchers just inconsistent, or are these measures not providing the whole picture?

3.2.4. Climax: How Do We Distinguish Between Skill and Luck in Pitching? As more scrutiny was given to traditional statistics, observers realized that it was not entirely fair to evaluate a pitcher solely on wins and ERA. As good as a pitcher may be, he cannot get a win by himself—his teammates have to score runs. Even though ERA does allow pitchers to avoid being penalized for defensive errors, it still depends a great deal on the quality of defense and on a good bit of luck. These facts led to a decades-long (and continuing) search for pitcher metrics that are independent of defense. Basco and Davies [3] provide a nice overview of the development of DIPS (defense-independent pitching statistics), and they tell the story of a researcher named Voros McCracken.

In the late 1990s, McCracken [57] made an effort to distinguish between things that pitchers can control and things that they cannot. In the former category were strikeouts, walks, and home runs—and he noticed that there was a fairly high level of consistency from year to year in these areas for individual pitchers. In the second category was a statistic called the batting average on balls in play (BABIP). This is essentially the percentage of non-home-run fair balls allowed that result in hits. McCracken [57] noticed that individual pitchers' BABIP numbers were not consistent at all from year to year. A recent example of this is Astros starting pitcher Zack Greinke. Greinke allowed a BABIP of 0.321 in 2020, after batters he faced hit just 271 on balls they put into the field in 2019 (Fan Graphs [30]). His strikeout, walk, and home run rates barely budged during this time (Fan Graphs [30]). What would cause the BABIP to be low one year and high the next? And why is this not uncommon? McCracken [57] suggested that this may represent a way to understand the luck factor.

As with many metrics that have been developed over the years, BABIP received varying levels of acceptance. McCracken's [57] premise was that pitchers have little control over their BABIP numbers. Others, like Tippett [84], argued (also with data) that this was not the case for every pitcher. While the debate over the usefulness of BABIP has continued, other pitcher metrics have been developed and analyzed as well. The fielding independent pitching metric tries to get at a pitcher's true skill by factoring in the things for which the pitcher has the most control: walks, strike outs, home runs, and hit by pitches. A pitcher's strand rate (left-on-base percentage) measures how well he leaves runners on base. These, and others, are reasonable approaches to evaluating pitchers, but as some analysts like Stoltz [81] have pointed out, they are not perfect.

3.2.5. Resolution: Luck Can Never Be Removed from Baseball. In 2015, Zack Greinke, then a Dodger, recorded 128 consecutive outs without surrendering a run (nearly 43 innings). Much was written about this remarkable streak, with many attempts to explain some of the reasons this could have occurred. Writing in the *Guardian*, Jack Moore [63] was critical of those who indicated that luck was a factor. His opinion was that writers should not be too quick to cite luck as a reason when it may be that the real reasons are simply unknown to us. Responding to Moore's criticism, writer Eric Garcia McKinley [31] says that "Luck can never be removed from baseball, and we'd be poorer if luck were removed from our analytical vocabulary about the game." He argues that luck is a reality in the game and that it makes it better. Analytical approaches to understanding luck do not intend to remove it from the discussion. Rather, these approaches attempt to help us all see the game with clearer lenses.

4. Prescriptive Analytics: Decision Making

Coaches and managers must consider game context in order to make good decisions and implement wise strategy. Prescriptive analytics models incorporate the score, remaining time in the contest, personnel, risk tolerance, and many additional factors to make better decisions. Coaches and general managers are often evaluated on the results of their decisions, but as many poker and business experts (Ma [54]) suggest, evaluation of decisions based on the results of those decisions is troubling because there are many aspects of the problem that are out of control of the coach. Analytics allows us to separate the strategy and the result for individual plays, games, and even whole seasons. In one famous example, during Super Bowl XLIX, Pete Carroll, head coach of the Seattle Seahawks, facing a second and goal from the opposing team's one-yard line and with 26 seconds left in the game, called a pass play instead of handing the ball to running back Marshawn Lynch. The result of the play was an interception by the opposing team, the New England Patriots. Fans and sports writers erupted at the outcome of the decision, stating that Marshawn Lynch (known at the time as Beast Mode for his ability to leave a mound of defender carnage in his wake when he ran the ball) could have scored from that distance easily. Some outlets (Schwartz [78]) came to the defense of Carroll's decision: "If it works, it was genius; if it didn't, it was a terrible play call." Other outlets (Morris [64]) cited statistics indicating that, during the year, teams who passed from the one-yard line had a touchdown success percentage of 60.9% versus 57.1% for runs. Furthermore, these outlets argued that even if the pass was incomplete on the second down, Carroll could have given Lynch a chance on the third down. Regardless, it is tough to judge the decision because both sides have a point. We do not know the actual decision-making process that Carroll employed. Whether it was based on the league percentages of passes versus runs from the one-yard line, the element of surprise, or something else, it is clear that decision making has an objective and is conditioned on the constraints of the current state of the game, and on the likelihood of outcome success based on the coach's sample of data.

Data visualization, like metrics for performance evaluation, tells a condensed story of the current state of a game. Consider heat maps produced in sports like darts (Lazzo [51]) and soccer (Robinson [74]). These visuals are a quick way to evaluate the in-game decisions of opponents as well as a player's own decisions. In darts, for example, the strategy of "going for it" is not always the best decision. Players accumulate points by hitting areas of the dart board worth varying amounts of points. Only aiming for the highest-valued areas slows the culmination of points by hitting less valuable targets more often. Like the aforementioned heat maps on shot selection in the NBA, dart players are now employing strategies based on their own historical heat maps and their opponents' heat maps to minimize the expected number of throws needed to reach 501 points, and intelligent betting on darts matches now makes use of historical player data (Lazzo [51]).

Golfers attempt to minimize the number of golf strokes needed to put the ball into the hole, and shooting for the flag may not be the best strategy. It is easy to visualize the precise golf hole location on a green as the analog to the dart board center bull's-eye. However, given the range of possibilities after a golf ball is struck toward an intended target, it is advantageous to use analytics and expected value to identify better, smarter targets for each hole. The presence of hazards, penalty shot areas, and difficult putting areas of a green make (sometimes barely) missing the intended target disastrous in golf by rapidly increasing the number of shots needed to finish the hole. In golf, embracing the strokes-gained statistic (Broadie [17]) has changed player preparation before and course management during golf tournaments by helping to identify better targets that minimize the expected number of strokes remaining to finish the hole.

How players and teams prepare and compete are fundamental components of stories about sports. Prescriptive analytics provides the mathematical tools to players, coaches, and fans to better understand and assess strategy and decision making before and during competition. Although the minutiae of expected value calculations and optimization frameworks are not necessary during in-game broadcasts, there is an important place for analytics to translate the risk–reward elements navigated by the protagonists of the stories. Major League Baseball shows a batter's frequency maps across sectors of the field and leads the viewer to question how the defense should react. The Professional Golf Association Tour television broadcasts include player's strokes-gained rankings by driving, approach, and putting categories to report how the golfer is playing in addition to the number on the scorecard.

4.1. Fair Play

An equal opportunity is fundamental to competition in sports. Paths to playoffs, travel schedules, and home field advantage must be balanced and distributed equitably among teams. Mathematical programming has proven highly effective as a tool to create sport league schedules, also called timetabling. Professional leagues (Goossens and Spieksma [34], Trick et al. [85]), collegiate leagues (Bouzarth et al. [10], Nemhauser and Trick [67]), and even recreational leagues (Grabau [35]) require fair schedules while incorporating dozens of team requests and league rules. Furthermore, within the structure of a season schedule, teams and fans demand fairness. Prescriptive analytics has emerged as an essential and impactful tool to create fair sport schedules and tell the story of each season.

4.1.1. Inciting Incident: Leagues That Have More Teams Than Regular Season Games Have the Problem That There Is No Head-to-Head Matchup Between Every Pair of Teams, Making the Problem of Deciding Who to Include in Playoff Structures Difficult. The NFL schedules regular season games each season so that a team plays its division foes twice (home and away), all teams from one division in each of the American Football Conference and National Football Conference (other than its own), and two games based on the previous year's results. In these last two games, which the authors call parity games, division winners from the previous season are matched up against other division winners, second-place teams in one division are matched up against second-place teams in another division, and so on. Because team composition changes from year to year, this scheduling policy creates variation in teams' strength of schedules and sometimes benefits teams unfairly, allowing some an easier path to the playoffs than others.

4.1.2. Progressive Complications: Making a Schedule to Balance All Metrics of Fairness Is Difficult. Scheduling could create unfairness on multiple criteria: travel distances, days off in between games, differences in quality of opponents. Easton et al. [28] cite many factors that influence scheduling fairness, including economic opportunities, television schedules, travel distances, balancing road trip duration with home stand duration, etc. Karwan et al. [43] identify and suggest how to fix the imbalance of opponent's rest days across NFL teams. Incorporating all of the fairness constraints could lead to models that are infeasible or take a long time to produce optimal solutions.

4.1.3. Crisis: Our Team Narrowly Missed the Playoffs This Year. How Much of That Was due to the Brutal Schedule That We Played? Table 3 shows the 2017 regular season results for five NFL teams that were in playoff contention. In the table, we use a quality measure and a strength of schedule for each team based on their rating percentage index (RPI). The RPI rating (Langville and Meyer [50]) for a team is defined as an weighted average of (a) a team's winning percentage, (b) the team's opponents' winning percentage, and (c) the teams' opponents' opponents' winning percentage. Terms (a) and (c) are equally weighted, and term (b) is weighted twice as much as (a) and (b). Note that Jacksonville and Tennessee made the playoffs but did so by playing a significantly weaker schedule (SOS scores of 45.17 and 45.59, respectively) than teams with similar win–loss records (Seattle, Detroit, and Dallas), highlighting that some teams get an easier path than others to the playoffs.

4.1.4. Climax: Bouzarth et al. [11] Introduce Methods to Produce an NFL Schedule That Eliminates These Games Based on the Previous Year's Results and Replaces Them with Two Games (One Home, One Away) Aimed at Reducing the Variability of Teams' Strengths of Schedule So That No Team Has an Easier Path to the Playoffs Than Another. The authors introduce an integer programming model to dynamically schedule two parity games at the end of the NFL season once the rest of the season has completed. The model uses a binary decision variable to indicate when two teams play a game in a given week of the season. The authors employ the expected change in NFL

Table 3. A comparison of 2017 playoff contenders.

Team	Record	Division record	SOS	RPI	Made playoffs?
Jacksonville	10–6	4–2	45.17	49.51	Yes
Tennessee	9–7	5–1	45.59	48.25	Yes
Seattle	9–7	4–2	50.00	51.56	No
Detroit	9–7	5–1	51.07	52.36	No
Dallas	9–7	5–1	50.82	52.18	No

schedule variability as an evaluation metric. The objective function minimizes the standard deviation of the strengths of schedule of all NFL teams. Constraints enforce requirements that each team plays once per week and that among the final two games, each team has one home and one away game against different teams. Details of the optimization framework are given by Bouzarth et al. [11].

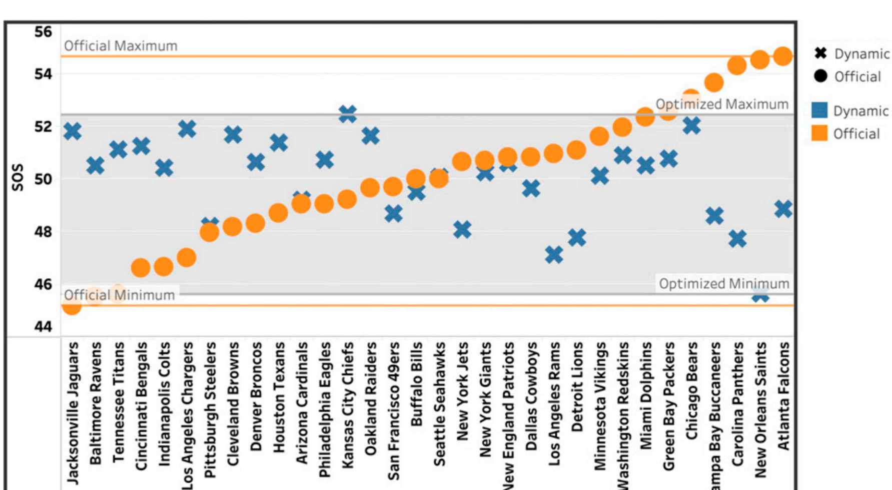
4.1.5. Resolution: The New Schedule Offers a 41.2% Reduction in the Variation of Strength of Schedules for the NFL Teams. Figure 4 shows the SOS numbers for teams scheduled with the official NFL schedule and for the schedule with dynamically scheduled parity games using the integer programming approach. The figure shows that the dynamically scheduled season exhibits a significant reduction in the variation in SOS measures for the teams. Figure 5 shows how the teams in Table 3 have their SOS values changed as a result of the dynamic scheduling. Note that both Jacksonville and Tennessee play harder schedules, whereas Detroit and Dallas play easier schedules. Seattle, which was closer to an SOS of 50 to begin with, sees only slight changes to its schedule.

4.2. Strategy

The world is stochastic, and sports are the perfect vehicle to highlight the notion that anything can happen and often does. Simulation is a tool used for understanding randomness and developing better strategies with applications in gambling, in-game management, draft preparation, player evaluation, and March Madness tournament analysis. Bouzarth et al. [9] use simulation to show the improvement in BABIP after optimizing the defensive positioning of fielders in Major League Baseball.

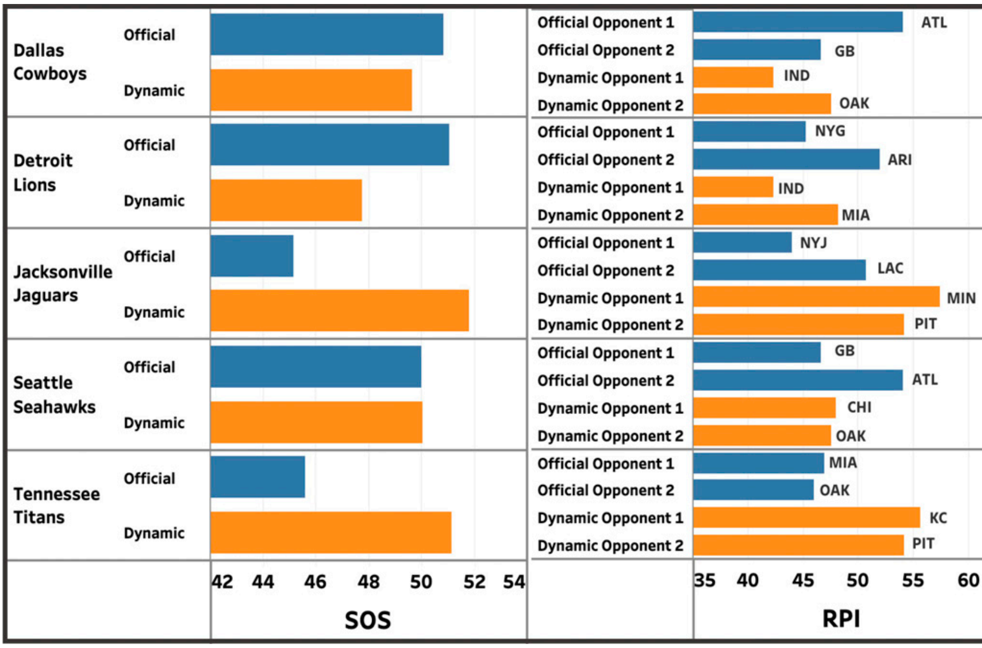
4.2.1. Inciting Incident: Defensive Repositioning Strategies (Shifts) Have Become More Prevalent in Major League Baseball in Recent Years. In 2018, batters faced some form of the shift in 34% of their plate appearances (Sawchick [76]). Traditionally, teams have employed a shift that overloads one side of the infield, especially against left-handers that have a tendency to pull the ball. However, in recent years, teams have experimented with novel ways of positioning players throughout the field in order to get batters out.

Figure 4. Comparison of the official and dynamic schedules for the 2017 NFL season.



Source. Adapted with permission from Bouzarth et al. [11], copyright @2020, IOS Press.

Figure 5. Simulated change in SOS for selected teams in 2017.

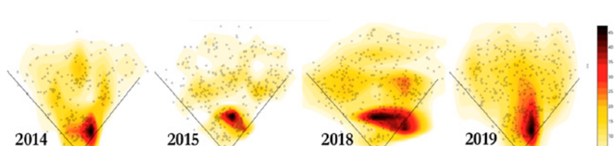


Source. Adapted with permission from Bouzarth et al. [11], copyright @2020, IOS Press.

4.2.2. Progressive Complications: Players Have Shown the Ability to Adjust to the Traditional Shift over Time Either by Trying to Lift the Ball or Hitting to the Opposite Field. Figure 6 shows Bryce Harper's hit distribution heat maps from balls in play from 2014 to 2019. As teams have employed a traditional shift against Harper, we can see a tendency for Harper to try to lift the ball into short center field (in 2018 and 2019), indicating an ability to change his hitting style to adjust to the defense.

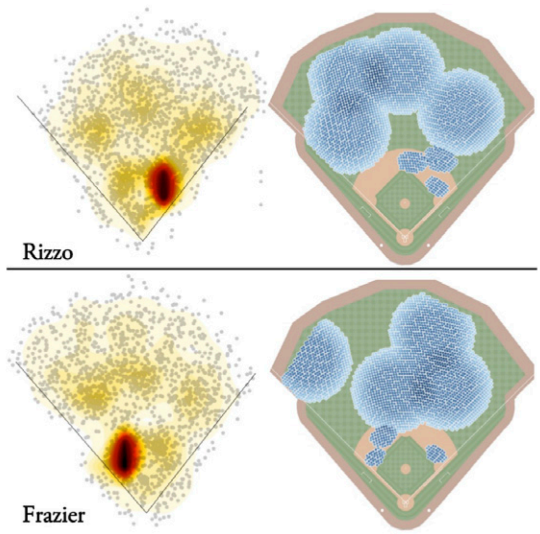
4.2.3. Crisis: Given That Hitters Are Meeting with Success Against a Shifted Defense, a Team's Infielders Need to Be Able to Adjust to a Batter's Changing Hit Distribution. A manager has to consider many possibilities in determining the positioning of players. The positioning should cover the areas of the field where the batter typically hits the ball, but at the same time, if the positioning leaves open areas of the field where the batter could hit the ball and achieve extra bases as a result, the manager would like to avoid this. These authors present a model that incorporates each of these in the objective function to balance the coverage of a batter's hit chart and the field's risk areas and contains a parameter allowing the manager to adjust the relative weighting of the two. The left side of Figure 7 shows

Figure 6. Heat maps of Bryce Harper's hit distribution from balls in play, 2014–2019.



Source. Copyright ©2021, INFORMS, Republished with permission of Walter de Gruyter and Company, from "Swing shift: A mathematical approach to defensive positioning in baseball," Bouzarth et al. [9], 17(1):47–55, 2021.

Figure 7. Shifted coverage of Anthony Rizzo and Todd Frazier.



Source. Copyright ©2021, INFORMS, Republished with permission of Walter de Gruyter and Company, from "Swing shift: A mathematical approach to defensive positioning in baseball," Bouzarth et al. [9], 17(1):47–55, 2021.

heat maps for two batters, left-handed Anthony Rizzo and right-handed Todd Frazier from the 2014–2018 MLB seasons. This type of heat map is incorporated into the model as a batter's intensity rating that goes into the objective function. The extreme areas of the field (the baselines and warning track) define the extrabase risk areas.

4.2.4. Climax: Bouzarth et al. [9] Describe a Prescriptive Analytics Approach to the Positioning of Players over the Entire Field of Play Without the Limitations of Traditional Positions or Current Methods of Shifting. The goal of the research was to position fielders in a way that maximized the coverage of a batter's intensity ratings and the risk areas. The model assumes the pitcher and catcher have fixed locations, whereas the constraints ensure that each player other than the pitcher and catcher is placed in the field so that fielders are sufficiently separated and no two defensive players are within a four-degree angle of each other relative to home plate. The constraints also ensure no area of the field is overcovered. A player assigned to a location can cover a range of locations around this placement with a score of one at this location that decreases to zero within a specified orb around the player as defined by how far an average player could reasonably move when reacting to a line drive. There is a constraint to place a player within a certain area around first base and optional constraints to place players within an area normally defined as the infield. The right images in Figure 7 show the shifted coverage against Anthony Rizzo and Todd Frazier as a result of the integer program.

4.2.5. Resolution: Was the Integer Programming Methodology Effective at Reducing Batters' Batting Average from Balls Put in Play? Using 2019 data obtained from baseballsavant.mlb.com, the authors created a hit distribution for a variety of left-handed and right-handed batters and simulated 10,000 balls in play to gauge the effectiveness of the positioning described by the integer programming model versus traditional placement and traditional shifted placement of defenders. The simulations show that an optimal positioning with three infielders lowered predicted BABIP by 5.9% for right-handers and by 10.3% for left-handers on average when compared with a standard four-infielder placement of players.

5. Getting Started and Some Lessons Learned

The use of analytics in sports is growing, and there is ample opportunity to be a part of the story. Conferences centered around sports provide a venue to bring together both sides of the storytelling team, with skills to analyze data and effectively present the results. The academic and industry partnership behind some of the work cited in this tutorial began with a conversation at the MIT Sloan Sports Analytics Conference. For the interested reader, joining the SpORts Section of INFORMS is a good place to start getting involved in sports analytics, and there exist lists (EURO OR in Sports [29], Sports Analytics Conferences [80]) of archived and upcoming sports analytics conferences, which are a great place to search for local, regional, and national networking opportunities.

Another path to successful storytelling with sports analytics goes through lower-level sports organizations. Professional franchises have normalized the contribution from all three types of analytics by way of dedicated analysts, teams, and entire departments. It has been our experience, however, that sports played outside the professional ranks provide many fruitful opportunities to conduct impactful and interesting research. Examples include minor leagues (still professional but without the same resources), collegiate sports, high school sports, and even recreational/community organized sports. These sports organizations are hungry for the impact of analytics but do not usually have the expertise on staff.

The current set of descriptive, predictive, and prescriptive analytics tools are not perfect, and there is a lot of work needed to bridge the gap between the analytical tools and perception among players, coaches, and the public. Skepticism around analytics exist because of the gulf between the analytics capabilities and the eye test. The use of analytics in baseball is abundant because there are individualistic aspects of the game that allow for player-specific metrics. Evaluative and predictive metrics for team-based sports have improved, but there is need for more, especially because the technology involved in these improvements and the data generated to analyze team play are proprietary. We hope that you will view this work as an invitation to join the sports analytics team.

We consider our own work as contributing to the evolution of a story in sports. So, we close this tutorial with lessons learned and some tips to consider when telling sports stories with analytics:

1. Your audience might not understand the role of randomness or variation in a result. Small sample sizes lead to large variation in results, especially when luck plays a big role. When announcers say that a batter is hitting 0.250 against a pitcher but that result is based on one hit in four attempts, it is difficult to determine whether the batter would naturally hit 0.200 or 0.300 against this pitcher.

2. Recognize the different objectives of different audiences. It is sometimes convenient to perform an analysis because it corresponds to the data at hand. Often the decision maker (coach, manager, etc.) has a different objective in mind. When presenting our own work in team evaluation to our school's basketball coaches, it was clear that the coach did not care that his team was ranked fifth in defense, for example. That number was not "coachable." The coach's objective is to create situations where his or her players learn and develop skill.

3. Answers are sensitive to assumptions. Question the assumptions that are being made with a model and perform sensitivity analysis to see how your answers change if the assumptions change. Our defensive positioning in baseball analysis had to be adjustable based on changes that batters make in their hit distribution.

4. Tell the full story. Determine what an adversarial opinion might be, and incorporate that argument into the story. Think of scenarios where someone's opinion would be supported by the data. What would have to be true? The changes we proposed in NFL scheduling introducing parity games to equalize strength of schedules had to acknowledge that it comes with a cost of potential greater travel and pushback from coaches in terms of preparing for uncertain opponents.

5. Ground your analytics story in an example. When the analysis starts to become complex and overwhelming, give the audience a concrete example to fall back on. In the Giant–Killer analysis, providing classic teams to associate with each cluster allowed the audience to connect to their past experiences watching the NCAA tournament.

6. The future of research may be to create complex models incorporating multiple operations at once such as analyzing the interdependence of concessions and player evaluation. Furthermore, because one model rarely incorporates all of the information needed to make evaluations, predictions, or decisions, consider building ensemble models that weight different viewpoints and metrics. As with many machine learning applications, though, story aspects are lost using these approaches because of the difficulty of explaining the results.

7. Last, the authors recommend to researchers starting out on sports analytics projects to learn how to scrape data from the internet. The lack of quality and open access data sets is problematic in this space but often fixable with a few lines of scraping code, especially when interested in high-frequency updates, that is, play-by-play data.

References

[1] AirNow. AQI basics. Accessed March 22, 2021, https://www.airnow.gov/aqi/aqi-basics/.

[2] K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy* 58(4): 328–346, 1950.

[3] D. Basco and M. Davies. The many flavors of DIPS: A history and an overview. *Baseball Research Journal* 39(2):41–50, 2010.

[4] Basketball Reference. 1979–80 NBA season summary. Accessed April 27, 2021, https://www.basketball-reference.com/leagues/NBA_1980.html.

[5] Basketball Reference. 2000–01 NBA season summary. Accessed April 27, 2021, https://www.basketball-reference.com/leagues/NBA_2001.html.

[6] Basketball Reference. 2020–21 NBA season summary. Accessed April 27, 2021, https://www.basketball-reference.com/leagues/NBA_2021.html.

[7] Basketball Reference. Michael Jordan. Accessed April 27, 2021, https://www.basketball-reference.com/players/j/jordami01.html.

[8] Basketball Reference. NBA single season leaders for value over replacement player. Accessed June 8, 2021, https://www.basketball-reference.com/leaders/vorp_season.html.

[9] E. Bouzarth, B. Grannan, J. Harris, A. Hartley, K. Hutson, and E. Morton. Swing shift: A mathematical approach to defensive positioning in baseball. *Journal of Quantitative Analysis in Sports* 17(1):47–55, 2021.

[10] E. L. Bouzarth, B. C. Grannan, J. M. Harris, and K. R. Hutson. Scheduling the Valley Baseball League. *INFORMS Journal of Applied Analytics*, forthcoming, 2021.

[11] E. L. Bouzarth, A. W. Cromer, W. J. Fravel, B. C. Grannan, and K. R. Hutson. Dynamically scheduling NFL games to reduce strength of schedule variability. *Journal of Sports Analytics* 6(4):281–293, 2020, https://doi.org/10.3233/JSA-200428.

[12] BracketOdds. University of Illinois. Accessed April 27, 2021, https://bracketodds.cs.illinois.edu/.

[13] J. C. Bradbury. Peak athletic performance and ageing: Evidence from baseball. *Journal of Sports Sciences* 27(6):599–610, 2009.

[14] J. Brenner and P. Keating. Upsets in the making. *ESPN The Magazine* (March 17), 17–22, 2014.

[15] J. Brenner and P. Keating. Upsets! We got upsets! *ESPN The Magazine* (March 2), 13–15, 2015.

[16] J. Brenner and P. Keating. Chaos kills. *ESPN The Magazine* (March 28), 20–22, 2016.

[17] M. Broadie. *Every Shot Counts: Using the Revolutionary Strokes Gained Approach to Improve Your Golf Performance and Strategy.* Avery, New York, 2014.

[18] B. Burke. Offenses run too often on first down. Advanced Football Analytics (November 23), http://archive.advancedfootballanalytics.com/2009/11/offenses-run-too-often-on-1st-down.html, 2009.

[19] T. P. Callaghan, P. J. Mucha, and M. A. Porter. Random walker ranking for NCAA Division IA football. *American Mathematical Monthly* 114(9):761–777, 2007.

[20] T. Chartier, E. Kreutzer, A. Langville, and K. Pedings. Bracketology: How can math help? J. A. Gallian, ed. *Mathematics and Sports*. Mathematical Association of America, Washington, DC, 55–70, 2010.

[21] J. Coleman. Minimizing game score variations in college football rankings. *Interfaces* 35(6): 483–496, 2005.

[22] College Football Playoff. Selection committee protocol. Accessed April 27, 2021, https://collegefootballplayoff.com/sports/2016/10/24/selection-committee-protocol.aspx, 2012.

[23] W. N. Colley. Colley's bias free college football ranking method: The Colley matrix explained. Accessed April 27, 2021, https://www.colleyrankings.com/matrate.pdf, 2002.

[24] D. Delen. *Prescriptive Analytics: The Final Frontier for Evidence-based Management and Optimal Decision Making.* Pearson, London, 2020.

[25] S. Deshpande and S. T. Jansen. Estimating an NBA player's impact on his team's chances of winning. *Journal of Quantitative Analysis in Sports* 12(2):51–72, 2016.

[26] D. Dilipkumar. Regression to the mean and its implications. Toward Data Science (March 24), https://towardsdatascience.com/regression-to-the-mean-and-its-implications-648660c9bf76?gi=61891bb248d0, 2020.

[27] S. Dutta, S. H. Jacobson, and J. J. Sauppe. Identifying NCAA tournament upsets using balance optimization subset selection. *Journal of Quantitative Analysis in Sports* 13(2):79–93, 2017.

[28] K. Easton, G. Nemhauser, and M. A. Trick. The traveling tournament problem description and benchmarks. *Proceedings of the Seventh International Conference on the Principles and Practices of Constraint Programming*. Lecture Notes in Computer Science, vol. 2239. Springer, Berlin, 580–584, 2001.

[29] EURO OR in Sports. Accessed April 29, 2021, https://www.euro-online.org/websites/orinsports/.

[30] Fan Graphs. Zack Greinke. Accessed April 27, 2021, https://www.fangraphs.com/players/zack-greinke/1943/stats?position=P.

[31] E. Garcia McKinley. Luck plays an undeniable role in baseball, and it has a place in baseball analysis. *SBNation* (August 19), https://www.beyondtheboxscore.com/2015/8/19/9175179/luck-plays-an-undeniable-role-in-baseball-and-it-has-a-place, 2015.

[32] K. Goldsberry. CourtVision: New visual and spatial analytics for the NBA. Presentation, *Sixth Annual MIT Sloan Sports Analytics Conference*, MIT Sloan, Cambridge, MA, 2012.

[33] K. Goldsberry. How mapping shots in the NBA changed it forever. FiveThirtyEight (May 2), https://vethirtyeight.com/features/how-mapping-shots-in-the-nba-changed-it-forever/, 2019.

[34] D. Goossens and F. Spieksma. Scheduling the Belgian soccer league. *Interfaces* 39(2):109–118, 2009.

[35] M. Grabau. Softball scheduling as easy as 1-2-3 (strikes you're out). *Interfaces* 42(3):310–319, 2012.

[36] C. K. Harrison and S. Bukstein. *Sports Business Analytics: Using Data to Increase Revenue and Improve Operational Efficiency.* CRC Press/Auerbach Publications, Boca Raton, FL, 2016.

[37] J. Hollinger. *Pro Basketball Prospectus.* Potomac, Dulles, VA, 2003.

[38] J. Hollinger. *Pro Basketball Forecast*, 2004–05 ed. Potomac Books, Dulles, VA, 2004.

[39] S. Ilardi and A. Barzilai. Adjusted plus–minus ratings: New and improved for 2007–2008. Accessed June 8, http://www.82games.com/ilardi2.htm, 2021.

[40] INFORMS. Operations research & analytics. Accessed March 16, https://www.informs.org/Explore/Operations-Research-Analytics, 2021.

[41] Investopedia. Cost per thousand. Accessed March 22, 2021, https://www.investopedia.com/terms/c/cpm.asp.

[42] B. James. *Bill James Baseball Abstract.* Free Press, New York, 1977.

[43] M. Karwan, M. Kurt, N. K. Pandey, and K. Cunningham. Alleviating competitive imbalance in NFL schedules: An integer-programming approach. Presentation, *Ninth Annual MIT Sloan Sports Analytics Conference*, March 5, MIT Sloan, Cambridge, MA, 2015.

[44] P. Keating. Play like a champion today? *ESPN The Magazine* (December 24), 9, 2012.

[45] P. Keating. The fall of Giants. *ESPN The Magazine* (March 27), 14–18, 2017.

[46] P. Keating. Crucial chaos. *ESPN The Magazine* (February 5), 8, 2018.

[47] A. Khatibi, D. M. King, and S. H. Jacobson. Modeling the winning seed distributions of the NCAA Division I Men's Basketball Tournament. *Omega* 50(1):141–148, 2015.

[48] C. N. Knaflic. *Storytelling with Data: A Data Visualization Guide for Business Professionals.* Wiley, Hoboken, NJ, 2015.

[49] P. Kvam and J. Sokol. A logistic regression/Markov chain model for NCAA basketball. *Naval Research Logistics* 53(8):788–803, 2006.

[50] A. M. Langville and C. D. Meyer. *Who's #1? The Science of Rating and Ranking.* Princeton University Press, Princeton, NJ, 2012.

[51] J. Lazzo. How to bet on darts: Trebles, big fish and more. *The Athletic* (May 27), https://theathletic.com/2390788/2021/05/27/how-to-bet-on-darts-trebles-big-sh-and-more/, 2021.

[52] M. Lewis. *Moneyball: The Art of Winning an Unfair Game.* W. W. Norton & Company, New York, 2004.

[53] J. Lunardi and D. Smale. *Bracketology: March Madness, College Basketball, and the Creation of a National Obsession.* Triumph Books, Chicago, 2021.

[54] J. Ma. *The House Advantage: Playing the Odds to Win Big in Business.* St. Martin's Press, New York, 2010.

[55] K. Massey. Statistical models applied to the rating of sports teams. Accessed March 6, https://masseyratings.com/theory/massey97.pdf, 2021.

[56] M. J. Mauboussin. *The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing.* Harvard Business Review Press, Cambridge, MA, 2012.

[57] V. McCracken. Pitching and defense: How much control do pitchers have. Baseball Prospectus (January 23), https://www.baseballprospectus.com/news/article/878/pitching-and-defense-how-much-control-do-hurlers-have/, 2001.

[58] F. McGuire. Chapter 1. *Defensive Basketball.* Prentice Hall, Upper Saddle River, NJ, 11–15, 1959.

[59] R. McKee. *Story: Substance, Structure, Style, and the Principles of Screenwriting.* Regan Books, New York, 1997.

[60] J. Meoli. Glossary of baseball's most useful advanced stats and how to use them. *Baltimore Sun* (February 28), https://www.baltimoresun.com/sports/orioles/bs-sp-orioles-analytics-explainer-20180227-story.html, 2018.

[61] E. D. Mills and H. D. Mills. *Player Win Averages.* A.S. Barnes, New York, 1970.

[62] R. Minton. *Sports Math.* CRC Press, Boca Raton, FL, 2017.

[63] J. Moore. "Being wrong is hard": Why we rely on luck to explain the unexplainable. *Guardian* (August 12), https://www.theguardian.com/sport/2015/aug/12/being-wrong-is-hard-why-we-rely-on-luck-to-explain-the-unexplainable, 2015.

[64] B. Morris. A head coach botched the end of the Super Bowl, and it wasn't Pete Carroll. FiveThirtyEight (February 2), https://vethirtyeight.com/features/a-head-coach-botched-the-end-of-the-super-bowl-and-it-wasnt-pete-carroll/, 2015.

[65] D. Myers. About box plus/minus (BPM). Accessed June 8, 2021, https://www.basketball-reference.com/about/bpm2.html, 2020.

[66] National Collegiate Athletic Association. Selections 101. Accessed March 22, 2021, https://www.ncaa.org/about/resources/media-center/mens-basketball-selections-101.

[67] G. Nemhauser and M. A. Trick. Scheduling a major college basketball conference. *Operations Research* 46(1):1–8, 1998.

[68] NIH National Heart, Lung, and Blood Institute. Calculate your body mass index. Accessed March 22, 2021, https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm.

[69] S. Nover. The rise of daily fantasy and sports betting has created an economy of its own. *Vox* (January 29), https://www.vox.com/2020/1/29/21112491/daily-fantasy-sports-betting-dfs-merch-analysis-weatherman, 2020.

[70] D. Oliver. *Basketball on Paper.* Potomac, Dulles, VA, 2003.

[71] K. Pomeroy. Accessed April 30, 2021, https://kenpom.com/.

[72] Pro Football Reference. NFL season by season passing. Accessed April 27, https://www.pro-football-reference.com/years/NFL/passing.htm, 2021.

[73] I. Rein, B. Shields and A. Grossman. *The Sports Strategist.* Oxford University Press, Oxford, UK, 45–46, 2015.

[74] J. Robinson. Hot and bothered: What does your favourite player's heatmap say about them? Dream Team (May 14), https://www.dreamteamfc.com/c/news-gossip/397578/favourite-player -heatmap-lionel-messi/, 2018.

[75] D. T. Rosenbaum. Measuring how NBA players help their teams win. 82 Games (April 30), https://www.82games.com/comm30.htm, 2004.

[76] T. Sawchik. Don't worry, MLB-hitters are killing the shift on their own. FiveThirtyEight (January 17), https://vethirtyeight.com/features/dont-worry-mlb-hitters-are-killing-the-shift-on-their-own/, 2019.

[77] A. Schatz. Football outsider basics. Accessed April 27, 2021, https://www.footballoutsiders.com/info/FO-basics.

[78] G. Schwartz. It's time to stop being mad at the Seahawks for throwing the ball at the 1-yard line in Super Bowl 49. *SBNation* (June 5), https://www.sbnation.com/n/2018/6/5/17426540/seahawks-patriots-super-bowl-49-malcolm-butler-interception-run-the-dang-ball, 2018.

[79] N. Silver. A better way to evaluate NBA defense. FiveThiryEight (July 9), https://vethirtyeight .com/features/a-better-way-to-evaluate-nba-defense/, 2019.

[80] Sports Analytics Conferences. Accessed March 22, 2021, sportsanalyticsconferences.com.

[81] N. Stoltz. Here's a thought: Explaining levels of pitching statistics. Bleacher Report (July 15), https://bleacherreport.com/articles/218125-heres-a-thought-explaining-levels-of-pitching-statistics, 2009.

[82] Team Rankings. NFL team fumble recovery percentage. Accessed April 27, 2021, https://www.teamrankings.com/n/stat/takeaway-fumble-recovery-pct.

[83] J. Thorn and P. Palmer. *The Hidden Game of Baseball.* Dolphin Books, New York, 150, 1995.

[84] T. Tippett. Can pitchers prevent hits on balls in play? Diamond Mind Baseball (July 21), https://diamond-mind.com/blogs/baseball-articles/77333188-can-pitchers-prevent-hits-on-balls-in-play, 2003.

[85] M. A. Trick, H. Yildiz, and T. Yunes. Scheduling Major League Baseball umpires and the traveling umpire problem. *Interfaces* 42(3):232–244, 2012.

[86] U.S. Bureau of Labor Statistics. Comparing the consumer price index with the gross domestic product price index and gross domestic product implicit price deflator. *Monthly Labor Review* (March 30), https://www.bls.gov/opub/mlr/2016/article/comparing-the-cpi-with-the-gdp-price-index-and-gdp-implicit-price-deator.htm, 2016.

[87] R. Wright. College football: Understanding how the BCS rankings work. Bleacher Report (November 16), https://bleacherreport.com/articles/943348-college-football-understanding-how-the -bcs-rankings-work, 2011.

[88] J. Zwerling. NBA draft may not prize mid-range game, but stardom awaits those who master it. Bleacher Report (June 2), https://bleacherreport.com/articles/2067958-nba-draft-may-not-prize-midrange-game-but-stardom-awaits-those-who-master-it, 2014.